

UNIVERSIDAD TECNOLOGICA METROPOLITANA  
DEPARTAMENTO DE GESTION DE INFORMACION  
ESCUELA DE BIBLIOTECOLOGIA  
CATEDRA: ANALISIS DE INFORMACION  
PROFESORA: SRA. TEXIA IGLESIAS MATURANA

DIAGRAMA DE FLUJO PARA INDIZAR CON UN TESAURO

WELLISCH, Hans. "A flow chart for indexing with a thesaurus". Journal of the ASIS. 23(3):185-194, May/June 1972. (Traducción de Textia Iglesias para fines docentes).

\*\*\*\*\*+

DIAGRAMA DE FLUJO PARA INDIZAR CON UN TESAURO

WELLISCH, Hans. "A flow chart for indexing with a thesaurus". Journal of the ASIS. 23(3):185-194, May/June 1972. (Traducción de Texia Iglesias para fines docentes).

INTRODUCCION.

La mayoría de las operaciones de indización, ya sea que se realicen para un servicio general de indización o individualmente en una biblioteca especializada o centro de información, se apoyan en un vocabulario controlado representado por un tesoro o por una lista de encabezamientos de materia.

Hoy en día, la construcción y mantención de tesauros ha alcanzado una etapa en que es posible utilizar sofisticados programas computacionales para generar el marco de diversos despliegues de descriptores y sus referencias cruzadas, una vez que la lista básica de descriptores ha sido elegida y sus términos, jerárquica y conceptualmente relacionados, han sido arreglados de acuerdo a cierto patrón deseado. (1, 2, 3, 4).

Se intenta que las altamente estructuradas y formalizadas "ayudas para la indización" resultantes, se utilicen siguiendo una estricta rutina. Posiblemente, una rutina que incluso pudiese automatizarse hasta cierto grado, como un corolario a aquellas partes del tesoro que han sido generadas automáticamente.

El diagrama de flujo que se incluye en la figura 1, es un intento de listar todos los pasos necesarios para una total y sistemática utilización de todas las posibilidades contenidas en un tesoro.

El diagrama está, en gran parte, diseñado con relación al Tesoro TEST (Thesaurus of Engineering and Scientific Terms, del Engineer's Joint Council).

Se eligió este tesoro por ser uno de los pocos de naturaleza general y por su gran uso en la indización de documentos científicos y técnicos producidos en las sociedades de ingenieros de los Estados Unidos.

El diagrama, sin embargo, debe ser aplicable a cualquier tesoro bien construido, así como a sus antecesoras las listas de encabezamientos de materia tradicionales, como la "Library of Congress Subject Headings" y su versión abreviada la "Sear's List of Subject Headings" (aunque éstas y otras listas similares carecen de ciertas características, como la distinción entre términos "específicos" y términos "relacionados", los que aparecen cubiertos indiscriminadamente por la referencia "véase además").

Es probable que la mayoría de los indizadores experimentados usen un tesoro, más o menos rutinariamente, en la forma descrita en el diagrama de flujo, pero no siempre siguen cada paso hasta su conclusión lógica y, a menudo, pierden términos de indización adecuados por no basar su elección en una exploración sistemática de todas las listas de términos disponibles - TG, TE, o TR - y sus subdivisiones. Para estos indizadores, una mirada ocasional a los pasos delineados en el diagrama de flujo, puede ser un útil procedimiento de chequeo, especialmente cuando la elección de los términos de indización es difícil o, en cierto modo, no satisfactoria o no exhaustiva del tema que trata el documento.

Para el indizador novato, el diagrama puede usarse como una ayuda de entrenamiento que lo introducirá en un uso más completo y exhaustivo de un tesoro. Después de haber indizado un cierto número de documentos con la ayuda de esta rutina, el uso del tesoro con todas sus ramificaciones llegará a ser casi un reflejo condicionado.

Esto no significa que el uso de un diagrama de flujo pueda hacer de la indización algo "automático" en el sentido de llegar a ser una mera aplicación mecánica de reglas a una lista dada de términos. Sólo intenta servir de guía en el uso de todos o la mayoría de los dispositivos disponibles en un buen tesoro, conteniendo además una rutina para los casos donde el tesoro no lista ningún término correcto o más adecuado, de entre un conjunto de términos candidatos posibles.

## EVALUACION DE UN DOCUMENTO PARA SER INDIZADO.

Antes de que el proceso de indización propiamente tal pueda empezar, son necesarios algunos pasos básicos. El documento a indizar debe ser revisado (scanned), vale decir, examinado de una manera general y en cierto modo superficial, a fin de establecer su elegibilidad para ser incluido en una colección de documentos y, por lo tanto, ser indizado para su posterior recuperación.

La expresión "De valor" en el primer rombo de decisión en el diagrama de flujo, no se refiere, por supuesto, al valor intrínseco del documento, sino sólo a su valor en relación a las necesidades de los posibles usuarios. //

Esta clase de "valor" tiene que ser asignado en forma diferente, dependiendo de si el documento va a ser indizado para un servicio general de índices y resúmenes dirigido a una gran clientela que tiene una amplia orientación temática en común, o si va a ser parte de una colección dirigida a un grupo especializado de usuarios con específicas necesidades de información, centrada en una o a lo más en unas pocas áreas temáticas bien definidas y específicas, como el caso de una biblioteca especializada o un centro de documentación orientado a la investigación.

Esta diferenciación afectará también, en definitiva, al tipo y número de términos de indización elegidos para el documento y tiene ciertas implicancias también, como se verá, en el uso del diagrama de flujo en varios puntos.

Un documento revisado para ser incluido, digamos, en el Engineering Index, será juzgado como valioso y por ende digno de ser indizado, si su contenido es de interés para la comunidad ingenieril en general y si su nivel de exposición es adecuado. Puede ser un artículo sobre "construcción de puentes", sobre "recuperación del óxido de zinc", sobre "propulsores de cohetes sólidos", o sobre cualquier tópico tecnológico, a condición de que tenga al menos alguna información nueva y esté escrito competentemente.

No es lo mismo en una colección especializada en Ingeniería Civil, por ejemplo, donde sólo se conservaría el artículo sobre "construcción de puentes", mientras que los otros dos serían "descartados" por no ser de valor para los usuarios de esa colección en particular.

La expresión "descartado" no implica que el documento sea en realidad destruido, sino simplemente que no se incluye en la operación de indización y que, tal vez, sea transferido a otra unidad de información que tenga que ver con el tema en cuestión y que, por lo tanto, pueda encontrar que el documento es "de valor" para sus propósitos.

## ELECCION DE LOS TERMINOS DE INDIZACION "CANDIDATOS".

Después de registrar los datos bibliográficos, que son indispensables para ubicarlo físicamente, el documento es leído y su contenido conceptual analizado. El texto del rectángulo etiquetado "Leer documento", necesita algo de extensión y explicación.

Primero que todo, el documento no necesita ser leído de "tapa a tapa" en cada instancia. Muchas veces puede venir acompañado del resumen del autor, y la lectura del título, sumario, introducción y conclusiones, serán suficientes para que el indizador se forme una opinión acerca de los conceptos tratados en el documento.

La instrucción "Subraye los TD que representan conceptos significativos", por supuesto no debe ser tomada literalmente. Aunque en la forma más rudimentaria y primitiva de la indización "unitérmino", el "indizador escribe una lista de palabras del texto que él piensa son representativas del contenido del documento" (5). El mero subrayado mecánico de sustantivos, adjetivos y verbos "significativos" o importantes, muy rara vez, si es que alguna, llevará a una indización exhaustiva de conceptos para una futura recuperación. La indización KWIC (que es otro método puramente mecánico), no puede lograr nada más que una simple indicación de palabras que a un autor se le ocurrió usar, sin revelar necesariamente todos o la mayoría de los conceptos tratados en el documento. El proceso de encontrar conceptos indizables y sus apropiados términos de indización, es bastante más complicado.

En la abundante literatura existente sobre este tópico, a lo menos hay una investigación, la de L.H. Oliver y otros (6), que analiza el proceso de indización en la forma de un diagrama de flujo, el cual puede substituirse por la versión condensada de las etapas preliminares del presente diagrama.

Una de las etapas finales en el diagrama de flujo de Oliver, es la operación etiquetada "Busque los descriptores para verificar que ellos son aceptables al sistema", y es justamente en este punto donde se centra nuestro análisis de los pasos necesarios en la indización.

No nos estamos preocupando del proceso mental que lleva a la formulación de un "resumen", pero presumiremos que alguien (que no necesariamente es siempre el indizador), realizó una primera etapa de condensación de las ideas y conceptos contenidos en el documento, dando como resultado un resumen, del cual se deriva una segunda etapa de condensación para propósitos de la indización, a través del subrayado de las palabras más significativas como posibles candidatos para ser incluidos en una lista final de términos de indización.

## USO DEL DIAGRAMA DE FLUJO: UN EJEMPLO.

Los términos de la "lista de términos candidatos", deben ahora ser revisados en el tesauro utilizado en el sistema, con la intención de indizar el documento dentro del marco aceptado de un lenguaje de indización prescrito.

Esto puede ser hecho a diferentes niveles de sofisticación, y aquí, otra vez, el factor "audiencia potencial" tendrá una influencia decisiva en la elección de los términos, aunque todos, o la mayoría de ellos, serán elegidos en definitiva de la lista establecida en el tesauro.

Un servicio de indización general tendrá que contentarse, indudablemente, con términos amplios, más generales (aunque no necesariamente menos). Por otra parte, un centro de información especializado, puede que tenga que usar algunos "términos más específicos", a veces no encontrados explícitamente en el documento o en su resumen, pero reconocidos por el indizador como útiles para la recuperación de ese documento por su clientela especializada,

En los ejemplos, se presume que la indización se está realizando para un centro de información que sirve a científicos e ingenieros, básicamente involucrados en la investigación y aplicaciones prácticas en un cierto campo temático específico.

El diagrama de flujo traza los pasos necesarios para la recuperación de todos o la mayoría de los términos potencialmente útiles de un tesauro en el que aparecen desplegados en varias configuraciones de relaciones jerárquicas o de correlación.

El punto focal es incluir un término en la lista como "término de indización" (TI). En este punto, todos los demás pasos quedan atrás, después de haber explorado varias posibilidades de encontrar tales términos de indización. Los diversos pasos seguidos serán mejor ilustrados por un ejemplo real.

Se eligió un artículo al azar del número de Diciembre de 1967 del Engineering Index, principalmente porque el Tesauro TEST *fué publicado ese mismo año* y era razonable presumir que su terminología coincidiría en gran medida con la usada en un artículo técnico publicado durante el mismo período.

El resumen del artículo (ver figura 2) se considerará como una adecuada primera etapa de condensación y las palabras más significativas para la indización y subsecuente recuperación.

En este ejemplo, naturalmente, la elección se hizo según el juicio del indizador, pero no parece que fueran posibles grandes variaciones cuando los términos de indización se extraen de un corto resumen indicativo, donde los términos normalmente se derivan del texto del documento mismo y, por lo tanto, son "términos del documento".

---

Computer Model of Slag Fuming Process for Recovery of Zinc Oxide. H.H. KELLOGG. Met. Soc. of AIME Trans. v.239, N°9, Sept. 1967. pp.1439-49.- ; / Model of slug fuming process for recovery of zinc oxide from lead blast furnace slags adapted to solution by digital computer, is presented: behavior of model is compared with operation of American Smelting and Refining Company's furnace, at El Paso, Texas: model is used to predict effects on performance of changes in coal rate, temperature of secondary air, and composition of coal.

---

Fig.2: Resumen con términos subrayados.

La figura 3 muestra un cuadro con los términos candidatos y los que, finalmente, quedaron como términos de indización,

| términos del documento (resumen) | Engineering Index | Términos de Indización (uso del tesoro) | Términos de Indización finales |
|----------------------------------|-------------------|---|--------------------------------|
| Blast furnace                    |                   | Blast furnaces                          | BLAST FURNACES                 |
| Coal                             |                   | Burning rates                           | Burning rates                  |
| Coal rates                       |                   | Coal gasification                       | Coal gasification              |
| D Composition                    |                   | Coal + rates                            |                                |
| e Computer model                 |                   | Composition                             |                                |
| s Digital computers              |                   | Computerized Simulation                 | COMPUTARIZED SIMULATION        |
| a                                |                   | Computers                               | / /                            |
| r Digital computers              |                   | Digital computers                       |                                |
| i Digital simulation             |                   | Digital simulation                      |                                |
| p Furnaces                       |                   | Furnaces                                |                                |
| t Heat of vaporization           |                   | Heat of vaporization                    | Heat of vaporization           |
| o Lead                           | Lead              | Lead                                    | LEAD                           |
| r Materials recovery             |                   | Materials recovery                      |                                |
| e Metallurgical furnaces         |                   | Metallurgical furnaces                  |                                |
| s Models                         |                   | Models                                  |                                |
| Performance                      |                   | Performance                             |                                |
| r Production models              |                   | Production models                       | Production models              |
| e Recovery                       |                   |   |                                |
| s Secondary air                  |                   |   |                                |
| Simulation                       |                   | Simulation                              |                                |
| Slag fuming                      |                   | Slags + Vaporizing                      | SLAGS + VAPORIZING             |
| Temperature                      | Smelting          |   |                                |
| Zinc oxide                       |                   | Thermal recovery methods                | Thermal recovery methods       |
|                                  |                   | Vaporizing                              | Vaporizing                     |
|                                  |                   | Zinc Oxides                             | ZINC OXIDES                    |
| i A.S.R.C.*                      |                   | A.S.R.C.                                | A.S.R.C.                       |
| d El Paso, Tex.                  |                   | El Paso, Tex.                           |                                |
| e U:S:                           |                   | U:S:                                    |                                |
| n.**                             |                   |   |                                |
| 15                               | 2                 | 25                                      | 12                             |

a) En el Índice Final los términos principales están en mayúsculas.  
 \* American Smelting and Refining Company  
 \*\* Identificadores.

Fig. 3: Tabla con los términos candidatos y finales.



Como se observa en la Figura 3, no todos los términos del documento se convierten en los mejores términos de indización, cuando se realiza el procedimiento de revisarlos en el tesoro y cuando se pone un límite al número de términos de indización. Mas tarde volveremos a este requisito restrictivo.

Ahora, los términos subrayados se comparan sistemáticamente con los del tesoro (\*) siguiendo las rutinas indicadas en el diagrama de flujo.

El primer término es COMPUTER MODEL. Este es un término complejo (es decir, consiste en más de un término, cada uno de los cuales puede existir por sí mismo para indicar un concepto, COMPUTADOR y MODELO), y no aparece listado como descriptor en el TEST; tampoco es un identificador.

Siendo así, el paso siguiente consiste en buscar un término sinónimo. Obviamente, una forma de hacer esto es buscar el término en un diccionario comprensivo o especializado donde puede que se usen sinónimos en la definición del término. En el caso de los términos complejos, a veces, es posible simplemente mirar uno de los términos componentes en el tesoro y buscar combinaciones de términos, que pueden ser sinónimos del término no listado.

Suponiendo que elijamos esta alternativa, comenzamos con COMPUTER (\*\*). Las primeras dos combinaciones, COMPUTER COMPONENTS y COMPUTER DRIVEN PUNCHES no son adecuados, pero el siguiente descriptor, COMPUTERIZED SIMULATION, parece ser el equivalente más cercano a "computer models" y, por lo tanto, se lista como TI.

El término tiene dos TG: MATHEMATICAL MODELS y SIMULATION, los cuales, en algunos sistemas, puede que fuesen listados automáticamente como TI, siguiendo la política de siempre registrar, además, los términos genéricos. Un ejemplo de tal política se puede encontrar en las prácticas de indización de EURATON (\*\*\*), donde para la mayoría de los términos de indización aparecen también consignados sus términos genéricos TG (7). Si se adopta tal política, hay que modificar el diagrama de flujo con una conexión directa entre la respuesta positiva a la decisión "¿Tiene el término TG?" y la respuesta "Anotar término como TI", pasando por alto la rutina evaluativa mostrada en el diagrama de flujo.

---

(\*) Sin considerar las diferencias morfológicas que no afectan el significado, como el singular y el plural. La forma usada en el tesoro debe seguirse.

(\*\*) N.del T. En el caso de los ejemplos, se ha considerado conveniente conservar los términos en inglés, por una parte, porque así aparecen en el TEST y, por otra parte, porque la forma de los descriptores compuestos traducidos al español, hace perder sentido al ejemplo.

(\*\*\*) Commission of the European Communities.

Es necesario destacar que la anotación automática de los TG puede hacerse solamente a condición de que todos los TG sean verdaderamente términos genéricos al descriptor, expresando un paso de supraordinación en una jerarquía basada en la relación género-especie. Una relación todo-parte, causa-efecto, o propósito-medio, nunca debe ser expresada como un TG, sino solamente como un TR.

Desafortunadamente, no todos los tesauros observan esta regla, y TEST no es una excepción. En nuestros ejemplos, todos los TG serán evaluados en relación a su propiedad y utilidad para la audiencia a que están dirigidos, siguiendo la rutina empleada para los TE y los TR, donde tal evaluación y elección debe hacerse inevitablemente. De los dos TG encontrados bajo COMPUTARIZED SIMULATION, para nuestros propósitos, sólo se retuvo SIMULATION.

Entre los TE de COMPUTARIZED SIMULATION, encontramos DIGITAL SIMULATION, la primera parte del cual aparece también en el resumen bajo "digital computer": por lo tanto, se lista como un posible TI.

Entre los TR está COMPUTERS precedido por un guión, lo que indica que este término tiene sus propios TG, TE y TR. Cuando miramos COMPUTERS, encontramos el TG DATA PROCESSING EQUIPMENT (descartado por no ser adecuado aquí) y un TE DIGITAL COMPUTERS, que corresponde al término en el resumen; el resto de los TE y TR no son adecuados.

Continuamos con SLAG FUMING, que es, nuevamente, un término complejo que no aparece listado y que, por lo tanto, se separa en sus partes, siguiendo la rutina indicada anteriormente. Se encuentra el término SLAGS que tiene como TR a METALLURGICAL FURNACE, un término que no está contenido explícitamente en el resumen pero que, aunque más bien general, parece aceptable como TI. También está marcado con un guión, resultando que en el listado su TG es FURNACES y uno de sus TE es BLAST FURNACES que también aparece en el resumen.

Cabe hacer notar que una característica especial del tesauro TEST es señalar los descriptores listados como TE o TR cuando éstos, a su vez, son miembros de una jerarquía, teniendo sus propios TG, TE y TR. (\*). Si esta característica se hubiese omitido (como sucede en mucho otros tesauros), el término BLAST FURNACES se hubiese encontrado más tarde subrayado, pero el término SIMULATION (genérico a COMPUTARIZED SIMULATION) se habría perdido.

---

(\*) Las explicaciones en el TEST establecen que un término precedido por un guión, denota la existencia sólo de TE pero, en realidad, la mayoría de los términos con guión tienen TG, TE y TR, cualquiera de los cuales podría ser útil como descriptor.

AMERICAN SMELTING AND REFINING COMPANY es un identificador, así como El Paso, Texas, al que puede agregársele U.S., ya que cualquier localidad también podría ser buscada por el estado o país en que está situado. En otras palabras, el TG de los identificadores (especialmente geográficos y taxonómicos) también deben listarse automáticamente.

PERFORMANCE ha sido subrayado y en realidad es un descriptor, pero su utilidad como TI es cuestionable debido a su generalidad y falta de contenido ("emptiness"): por el momento se lista como un posible candidato a TI final.

COAL RATE no se encontró como descriptor y otra vez, es necesario buscar una posible combinación COAL + RATE(S), (\*). En realidad, ambos pueden encontrarse como descriptores, y bajo RATES (PER TIME) encontramos un TE adecuado: BURNING RATE. Si esta combinación de términos se va a usar en el futuro como un TI y un término de búsqueda, nuevamente el indizador debe hacer una referencia USE en su ejemplar de trabajo del tesoro.

Coal Rate  
USE: COAL + RATES (PER TIME)

La inclusión de esta referencia USE en ediciones posteriores del tesoro impreso, es una política editorial que no nos compete aquí. Lo que sí es importante, es el registro sistemático de las decisiones de indización no cubiertas por el tesoro, así como las referencias resultantes USE y su recíproco UP (usado por).

El término TEMPERATURE, aunque un descriptor válido, es demasiado general, pero trae un TR HEAT OF VAPORIZATION que ajusta con el término VAPORIZING obtenido anteriormente. (Aquí hay una inconsistencia en el Tesoro TEST que debemos hacer notar - una de las muchas negligencias terminológicas de las que desafortunadamente adolece este tesoro. Si el término HEAT OF VAPORIZATION es un descriptor válido, ¿por qué no se permite "Vaporization" y en su lugar hay que usar VAPORIZING, que no es exactamente lo mismo?).

SECONDARY AIR no es un descriptor en el TEST, ni puede expresarse por la combinación de otros descriptores. Las operaciones y decisiones indicadas en el lado derecho del diagrama de flujo, entran ahora en juego.

---

(\*) La explicación (PER TIME) es necesaria ya que existe el término homónimo RATES (COSTS).

Una búsqueda de posibles sinónimos en un diccionario de metalurgia, no lleva a ningún sinónimo adecuado. Sin embargo, podría haber algunos términos genéricos adecuados en el Índice Jerárquico, donde los descriptores están agrupados en un patrón clasificatorio, ya que cualquier término puede aparecer de algún modo subordinado a un término jerárquicamente más amplio. Por esta razón, el casillero dice "Buscar... un posible TG al término", pues ahora ya no es cuestión de encontrar un TE o un TR; pero, cualquier término encontrado en una jerarquía clasificatoria debe ser un término más amplio o genérico, aunque un término encontrado en el más bajo nivel posible por sobre los TD no listados, podría aún ser demasiado general, como en realidad es aquí el caso.

"Secondary Air" es una característica de "Furnaces", término que, al mirarlo en el Índice Jerárquico, nos da los términos más específicos METALLURGICAL FURNACES y BLAST FURNACES (ambos ya anotados como TI), pero no otros términos bajos los cuales fuese posible agregar "secondary air". La cuestión, entonces, permanece: debe o no ser listado SECONDARY AIR como TI. Una decisión positiva podría basarse en: a) el juicio del indizador, que está basado a su vez en su conocimiento de los hábitos de recuperación de los usuarios, como, por ejemplo, su uso de la terminología; o b) la frecuencia con que ese término aparece en otros documentos relacionados.

El último criterio es particularmente aplicable a términos acuñados recientemente, cuya actualidad aún se desconoce. Se toma la decisión (que no se muestra en el diagrama) de mantener el término pendiente, en espera de nuevas apariciones en futuros documentos hasta un cierto número umbral, digamos diez. Luego que el nuevo término ha aparecido diez veces en diversos documentos, es adoptado como un TI válido e incorporado a la copia de trabajo del tesoro. Si el tesoro es puesto al día regularmente (principalmente como resultado de los hallazgos de varios indizadores mientras usan la última edición), debe hacerse una nota de inclusión del nuevo término para que sea considerado por los editores en la siguiente edición del tesoro.

En nuestro ejemplo, se presume que el término "Secondary Air" no sería un término buscado, por lo que el procedimiento de indización continúa con COMPOSITION. Este término aparece como descriptor COMPOSITION {PROPERTY} sin TE o TR apropiados, pero su adopción para indizar nuestro documento es cuestionable debido a su generalidad. Si se usase, debería estar ligado (linked) con COAL, el último término subrayado, ya listado como TI en otra combinación de términos. Finalmente, este término finaliza la lista de términos subrayados a ser revisados en el tesoro.

## ELECCION DE LOS TERMINOS DE INDIZACION FINALES.

Los resultados de este procedimiento de indización, tomando ventaja de todas las posibilidades ofrecidas por el tesauro, aparecen tabuladas en la columna 3 de la Figura N° 3. Los 16 TD originales subrayados, aumentado en 10 términos adicionales, 9 de los cuales son descriptores y uno es identificador.

En la mayoría de los sistemas de recuperación, 25 sería una cantidad excesivamente alta de términos, lo que a su vez podría influir en una alta incidencia de falsas recuperaciones, aun cuando las correlaciones indeseadas se hubiesen eliminado a través de adecuadas uniones (linking).

Es una práctica establecida, en gran parte de las operaciones de indización, el limitar el número de términos a usar, como asimismo el especificar un cierto número de términos principales o primarios, que aparecen en los índices impresos, y un número de términos secundarios almacenados en la base de datos sólo para propósitos de búsquedas por computador, y no disponibles para la búsqueda manual. Un ejemplo bien conocido de esta práctica es el MEDLARS.

Presumiendo que el sistema de recuperación fije un límite arbitrario de 12 términos de indización, 5 de los cuales pueden ser primarios, es necesario hacer una elección de entre los 25 términos surgidos del procedimiento antes descrito. Para los propósitos de esta discusión, se presume que la indización se realiza para un centro de información especializado, que sirve, fundamentalmente, a metalúrgicos interesados en especial en los metales no-ferrosos. Los términos de indización finales que se listan en la columna 4 de la Figura N° 3, han sido elegidos con estos parámetros en mente.

No se pretende aquí que la lista final de TI representa la mejor o la única posible. Sin embargo, lo que puede destacarse es que la mayoría de los términos logrados con el uso sistemático del tesauro son más específicos que los encontrados en el resumen. Aparte de los términos BLAST FURNACES, LEAD y ZINC OXIDES (dos de los cuales como se dijo son básicamente identificadores), el resto son todos más precisos: COMPUTARIZED SIMULATION es mejor que "Computer models" en cuanto a ambigüedad ("computer model" puede ser un "modelo de computador" y no un "modelo hecho con un computador" que sí implica "Computarized simulation"); COAL GASIFICATION es más preciso que COAL o COAL RATE; THERMAL RECOVERY METHODS es más preciso que simplemente RECOVERY, etc.

En el momento en que apareció el artículo, el Engineering Index seguía la política de indizar cada artículo bajo un solo término (a veces modificado por un subencabezamiento). Por cierto sería incorrecto comparar tal

práctica con una indización exhaustiva hecha con la ayuda de un tesoro, pero es interesante notar que el término SMELTING usado como encabezamiento en el Engineering Index (ver columna 2 en la Figura N° 3), no aparece en el resumen (excepto como parte del nombre de la Compañía), ni se encontró en el tesoro como un término asociado a ninguno de los términos de indización del resumen. Que un artículo sobre la "recuperación de zinc" deba ser indizado bajo "plomo" (lead), es francamente cuestionable, por decir lo menos.

#### OTROS EJEMPLOS.

Las figuras 4, 5, 6 y 7 muestran los resultados de la aplicación de los procedimientos del diagrama de flujo a otros dos artículos, ordenados en la misma forma que el ejemplo anterior (resúmenes y lista de términos). Para estos ejemplos también se considerará una clientela especializada para la confección del listado final de términos.

Para el artículo sobre "Construcción de puentes" (figuras 4 y 5), los resultados son similares a los mostrados en el primer ejemplo. Un ejemplo, TRAPEZOIDAL, no se encontró en el TEST y fué agregado: probablemente debería estar conectado (linked) a BOX BEAMS.

---

Bridge Construction at Penrith. Railway Gazette, v.123, n° 9, May 5, 1967. p.345-8. Trapezoidal section concrete girders slide into position so that the road construction can take the place beneath; abutments and piers were constructed under running conditions by sheet pile wall method; bridge girder is hollow post-stressed reinforced concrete box beam of trapezoidal cross-section, with central longitudinal stiffening member; advantages of sliding over rolling method are that greater hearing pressures can be permitted on sliding surfaces; sections of three bridges are shown.

---

Fig.4: Resumen con términos subrayados.

| col.1  | col.2                   | col.3   | col.4   |
|--|-------------------------|---|---|
| términos del documento (resumen)   | Engineering Index       | Términos de Indización (uso del tesoro)   | Términos de Indización finales  |
| Bearing pressures<br>Box beams<br>Bridge<br>Abutments<br><br>Concrete  | Bridges<br><br>Concrete | Beams<br>Bearing properties<br>Box Beams<br>Bridges<br>Bridge abutments<br>Bridge piers<br>Concrete construction<br>Concretes<br>Construction<br>Footings<br>Girder bridges<br>Girders<br>Highway bridges<br>Highways<br>Piers<br>Post tensioning<br>Pressure<br>Prestressed concrete<br>Reinforced concrete<br>Roads<br>Sheet piling<br>Sliding friction<br>Steel piles<br>Structural Engineering<br>Structural members<br>Trapezoidal | BOX BEAMS<br>Bridge abutments<br>Bridge piers<br>CONCRETE CONSTRUCTION<br><br>GIRDER BRIDGES<br>Highway Bridges<br><br>POST TENSIONING<br><br>Prestressed concrete<br><br>Sheet piling<br><br>Trapezoidal |
| D<br>Construction<br><br>e<br>Girders<br><br>s<br>Piers<br>r Post-stressed<br><br>i<br>p Reinforced concrete<br>t Road<br>o Sheet pile<br>Sliding<br><br>r<br><br>e<br>s Trapezoidal |                         |   |   |
| i Penrith<br>d<br>e<br>n   | Gt. Britain             | Gt. Brit.<br>Penrith  | Gt. Britain<br>Penrith  |
| 15   | 3                       | 28  | 12  |

Fig. 5: Tabla con los términos candidatos y finales.

En el caso del artículo sobre "Propulsores de cohetes sólidos" (figuras 6 y 7), los términos obtenidos al seguir el procedimiento sistemático son, en términos generales, no más específicos que aquellos obtenidos del resumen. Esto se debe, probablemente, a que el campo temático de los cohetes tiene un vocabulario altamente especializado y a que el tesoro TEST está particularmente sesgado hacia el tema de los cohetes y de los misiles, los que aparecen cubiertos con mayor amplitud y profundidad que otras ramas de la ciencia y la tecnología (debido al rol que juegan los expertos del Departamento de Defensa en las decisiones sobre términos a incluir en el TEST, el cual se basa en gran medida en el antiguo Proyecto LEX del Departamento de Defensa).

En este caso, el seguimiento de la rutina del diagrama de flujo lleva a resultados que son más bien opuestos, a los logrados en los otros dos artículos.

Algunos descriptores, como SOLID ROCKET FUELS y SOLID ROCKET PROPELLANTS son francamente redundantes. También hay descriptores tipo frase como SOLID PROPELLED ROCKET ENGINES, que podrían resultar de un valor cuestionable para la recuperación, pudiendo ser más conveniente dividirlo en constituyentes más básicos, usando los descriptores SOLID ROCKET PROPELLANTS y ROCKET ENGINES que aparecen en el TEST.

El término LOW PRESSURE RESEARCH debería reducirse a LOW PRESSURE (como se muestra en la columna 4 de la figura N° 7), ya que RESEARCH puede agregarse a casi cualquier tema que sea científicamente investigado. (\*)

Un indizador podría desear ésta y otras instancias similares, con miras a una futura edición del tesoro, con la intención de eliminar términos que tienden a dispersar la información sobre el mismo tema bajo dos o más descriptores diferentes, pero sinónimos o cuasi-sinónimos.

En otras palabras, el seguimiento de una serie de rutinas en el uso de un tesoro, puede conducir no sólo a adiciones necesarias, sino también al descarte de descriptores indeseados o innecesarios o de descriptores que han caído en desuso (un ejemplo reciente sería la substitución "Laser" por "Optical Maser"), y a un ajuste general de la terminología (como en el caso de "low pressure" antes citado).

---

(\*) Por alguna razón no explicada, RESEARCH se ha agregado en el TEST sólo a unos pocos temas como COAST GUARD RESEARCH (aunque no es muy claro, este descriptor se usa para investigaciones hecha por la Guardia Costera o para investigaciones sobre la Guardia Costera), INFRARED RESEARCH, MINING RESEARCH, etc., pero no Medical Research, Chemical Research u otros. Aparentemente no hubo una pauta clara para la inclusión del término RESEARCH en un descriptor complejo.



Low Pressure Combustion and Ignition of Solid Rocket Propellants. R, SERGAL, Astronautical Acta. v.13, N° 2, Mar.-April 1967, p.167-81. Design and application of low pressure solid rocket motor offers some advantages and poses some problems: later are associated with completeness of metal combustion, LP combustion instability, chuffing, ignition difficulties and other factors which result in performance degradation; results of theoretical and experimental investigations are presented; advantages which are not readily apparent (such as thrust termination, restart capability and reduction in two-phase flow losses with gas-particle system) are indicated; significance of low chamber pressure on performance, combustion and ignition of rocket motors is shown.

Fig. 6: Resumen con términos subrayados.

| col.1  | col.2             | col.3  | col.4   |
|--|-------------------|--|---|
| términos del documento (resumen)   | Engineering Index | Términos de Indización (uso del tesoro)  | Términos de Indización finales  |
| Chuffing<br>Combustion<br>Chamber<br><br>Combustion instability<br>Degradation<br>Gas-particle flow<br>Ignition<br>Losses<br>Low pressure<br>Metal combustion<br>Performance<br>Reduction<br>Restart capacity<br><br>Rocket motors<br><br>Solid rocket propellants<br>Two-phase flow<br>Thrust termination | Combustion        | Chuffing<br>Combustion<br>Combustion chamber<br>Combustion stability<br>Degradation<br>Gas flow<br>Ignition<br>Losses<br>Low pressure<br>Metal + Combustion<br>Performance<br>Reduction<br>Restartable rocket engines<br>Rocket engines<br>Rocket propellants<br>Solid propellants<br>Solid rocket fuels<br>Solid rocket propellants<br>Two-phase flow<br>Thrust termination | COMBUSTION<br>Combustion chamber<br>Combustion stability<br><br>IGNITION<br>Losses<br>LOW PRESSURE<br><br>Restartable rocket engines<br><br>SOLID ROCKET PROPELLANTS<br>Two-phase flow<br>Thrust terminat |
| 17   | 1                 | 21   | 10  |

Fig. 7: Tabla con términos candidatos y finales.

## USO DE DIVERSAS PARTES DEL TESAURO.

Aunque, a primera vista, podría parecer que el procedimiento esbozado en el diagrama de flujo es complicado y consumidor de tiempo, en realidad la mayoría de los indizadores experimentados utilizan el tesoro de esta forma, repitiendo en forma inconsciente todas o la mayoría de las rutinas para cada término de indización. El uso del diagrama de flujo como ayuda en la indización, asegurará la utilización consistente del tesoro y de todas o la mayoría de sus partes.

Se ha cuestionado si los diversos listados (displays) de descriptores que suelen incluir los tesoros son realmente útiles en el proceso de indización, o si en realidad son sólo la consecuencia de la facilidad con que tales listados pueden generarse por computador, una vez que los descriptores básicos y sus correspondientes TG, TE y TR han sido almacenados en la máquina (8).

El diagrama de flujo muestra que, al menos en el caso del TEST, todas las formas de listados pueden ser útiles aunque, por supuesto, el Listado Alfabético de Descriptores es el más usado.

El Índice Jerárquico es útil cuando uno se encuentra con términos no listados en el tesoro. El Índice Permutado guía al indizador a descriptores complejos en los cuales el término en que se pensó no es el primer componente (por ejemplo, el término COMBUSTION CHAMBERS en la Figura 7, se encontró de esta forma, ya que la nota de alcance bajo "Chambers" recomendaba el uso de un término más específico, y el término apropiado se encontró al buscar combinaciones de la palabra "chambers" en el Índice Permutado). Este índice presenta también ventajas para los usuarios de un sistema de recuperación de información, aunque no hay evidencia de que ellos realmente usen este tipo de listado cuando acuden a un servicio de información.

El Índice de Categorías Temáticas, que en el caso del TEST están en gran medida ordenadas por las categorías temáticas de COSATI (\*), probablemente sólo es útil para usuarios acostumbrados a este particular arreglo de las materias, ya que las categorías se escogen arbitrariamente y reflejan los intereses de una clientela especializada. No está dentro del alcance de este trabajo una discusión más detallada respecto a este punto.

---

(\*) COSATI = Committee on Scientific and Technical Information (este Comité pertenece al Consejo Federal de Ciencia y Tecnología de EE.UU.)

## CONSIDERACIONES DE IMPLEMENTACION

El diagrama de flujo ha sido construido como una ayuda sistemática en el uso de un tesoro para indizadores humanos. ¿Podría utilizarse también para la indización automática por computador? Desafortunadamente, la respuesta es no. Al menos no en el futuro inmediato. Si un computador pudiera asignar términos de indización a los documentos, siguiendo este relativamente simple diagrama de flujo, tendríamos aquí "el" descubrimiento para la indización, pero no es el caso.

Aunque la mayoría de las decisiones y operaciones en el diagrama podrían, en realidad, ser realizadas por un computador adecuadamente programado, en el cual hayan sido almacenados los términos del tesoro en una forma legible por la máquina, hay varios puntos en los que es necesario el juicio humano.

El primer punto, obviamente, es la decisión respecto al valor del documento para la colección, pero podríamos descartarlo suponiendo que tal decisión fué correctamente tomada por una mente humana en una etapa previa a la indización.

La decisión crucial, basada en un juicio valorativo, hay que buscarla en respuesta a la pregunta: "¿Hay algún término en la lista adecuado como TI?" Como hemos visto, el tesoro puede incluir TE y TR que no están expresamente mencionados en el documento o en el resumen, pero que pueden ser perfectamente adecuados como términos de indización. Ni los burdos métodos de indización KWIC, ni las técnicas más sofisticadas de relación o agrupación de términos, son en este momento capaces de realizar tal tarea con un nivel de éxito ni remotamente similar al que puede alcanzar un buen indizador.

La rutina de buscar términos, no encontrados en el tesoro, usando diccionarios de sinónimos, o co-ocurrencia estadísticamente válida en el documento o en documentos indizados previamente bajo uno o más de los términos asociados, aún podría dar como resultado muy pocos o demasiados (tal vez no-relevantes) términos candidatos. La situación es similar cuando debe revisarse un término no listado en un índice jerárquico en busca de posibles TG. Tales términos no pertenecen todos a una misma jerarquía y, aún si así fuese, la decisión sobre un TG como substituto aproximado del término no listado, también debe basarse en un juicio valorativo, como se muestra en los ejemplos que hemos visto.

El tercer punto en que se necesita el juicio humano, es aquél en que es necesario decidir si se admite o no como término de indización un término no listado. Aquí, sin embargo, existe una posibilidad, al menos parcial, de programar el

computador de tal manera que un término no-listado sea automáticamente rechazado como TI, pero almacenado y, posteriormente, admitido como TI si, y sólo si, se ha alcanzado un cierto número umbral de ocurrencias.

Por lo tanto, pareciera que, aunque las rutinas computacionales pueden hoy en día ayudar en la construcción de un tesoro, su uso depende en gran medida de decisiones intelectuales de seres humanos.

Por otra parte, esto no debe interpretarse en el sentido de que sólo expertos altamente entrenados pueden usar un tesoro en forma satisfactoria. En una reunión reciente sobre tesoros estructurados (9), se informó que indizadores trabajando con un buen tesoro y relativamente conocedores del área temática de la colección, podrían realizar el 80% de su trabajo sin necesidad de consultar a expertos profesionales, de manera que gran parte del trabajo "casi podía ser delegado a personal ayudante", manejando la operación de indización en base al principio de "administración por excepción". El diagrama de flujo, susceptible de adaptarse a las características especiales del tesoro en uso en cualquier sistema, puede ser de ayuda en tales prácticas de indización.

#### REFERENCIAS.

1. BLAGDEN, J.F. "Thesaurus compilation methods: a literature review". *Aslib Proceedings*. 20:345-359, August 1968,
2. HERSEY, D.F. and HAMMOND, W. "Computer usage in the development of a water resources thesaurus". *American Documentation*. 18(4):209-215, October 1967.
3. SESSIONS, V. and SLOAN, L.W. *Urbandoc, a bibliographic information system. Technical Supplement 2: Operation manual*. New York, City University of New York, 1971.
4. HINES, T.C. and HARRIS, J.L. "Columbia University School of Library Service System for Thesaurus Development and Maintenance". *Information Storage and Retrieval*, 7(1):39-50, June 1971.
5. BOURNE, C.P. *Methods of Information Handling*. New York, Wiley, 1963. p. 15.
6. OLIVER, L.H. et al. *An Investigation of the Basic Processes Involved in the Manual Indexing of Scientific Documents*. Bethesda, MD, General Electric Co. 1966.

7. ROLLING, L.N. "Compilation of thesauri for use in computer systems". Information, Storage and Retrieval. 6(4):341-350, June 1970.
8. SURACE, C.J. The Displays of a Thesaurus. Santa Monica, Rand Corp., 1970.
9. BLAGDEN, J.F. "Structured thesauri. Aslib Proceedings. 23(3):139-143, March 1971.

\*~\*\*\*\*\*

6 6

1



## SIMBOLOGIA USADA EN EL DIAGRAMA DE FLUJO

|                            |   |
|----------------------------|---|
| DESCRIPTOR (descriptor)    | = Término autorizado listado en el Tesauro.   |
| IDENTIFIER (identificador) | = Término autorizado no incluido en el Tesauro, como nombres propios, marcas comerciales, de lugares geográficos, nombres taxonómicos, etc. |
| TERM (término)             | = Cualquier palabra o frase usada para denotar un concepto., ,  |
| BT (broader term)          | = TG (término genérico)   |
| DT (document term)         | = TD (término del documento); término utilizado en el texto del documento.  |
| IT (index term)            | = TI (término de indización); término bajo el cual el documento es indizado.  |
| NT (narrower term)         | = TE (término específico)   |
| RT (related term)          | = TR (término relacionado)  |
| UF                         | = UP (usado por)  |